



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2016년08월19일
 (11) 등록번호 10-1649146
 (24) 등록일자 2016년08월11일

(51) 국제특허분류(Int. Cl.)
 G06F 17/30 (2006.01)
 (52) CPC특허분류
 G06F 17/30011 (2013.01)
 (21) 출원번호 10-2015-0007547
 (22) 출원일자 2015년01월15일
 심사청구일자 2015년01월15일
 (65) 공개번호 10-2016-0088503
 (43) 공개일자 2016년07월26일
 (56) 선행기술조사문헌
 KR1020110009744 A*
 KR1020110135226 A*
 KR1020120030389 A*
 KR1020110024948 A
 *는 심사관에 의하여 인용된 문헌

(73) 특허권자
 주식회사 카카오
 제주특별자치도 제주시 첨단로 242(영평동)
 (72) 발명자
 강지훈
 제주특별자치도 제주시 과원로 11 202동 101호
 (노형동, 부영)
 (74) 대리인
 권영준

전체 청구항 수 : 총 14 항

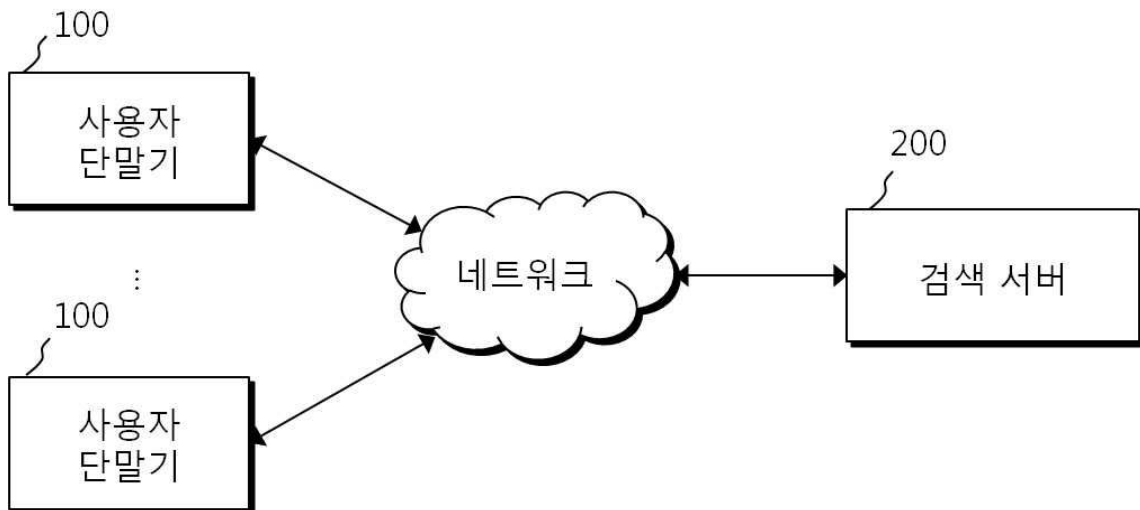
심사관 : 경연정

(54) 발명의 명칭 검색 방법 및 검색 서버

(57) 요약

검색 방법 및 검색 서버가 제공된다. 상기 검색 방법은 사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 단계, 상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에 관련된 복수의 웹 문서를 검색하는 단계, 각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 단계, 및 상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 단계를 포함하고, 상기 문서 순위 점수를 산출하는 단계는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하는 단계를 포함하는 것을 특징으로 한다.

대표도 - 도1



명세서

청구범위

청구항 1

사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 단계;

상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에 관련된 복수의 웹 문서를 검색하는 단계;

각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 단계; 및

상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 단계를 포함하고,

상기 문서 순위 점수를 산출하는 단계는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하는 단계를 포함하며,

상기 복수의 모수 중 적어도 하나를 업데이트하는 단계는 상기 복수의 웹 문서의 과거 문서 순위 점수와 검색 만족도의 오차를 이용하여 상기 복수의 모수 중 적어도 하나를 업데이트하는 것을 특징으로 하는, 검색서버에서 수행되는 검색 방법.

청구항 2

사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 단계;

상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에 관련된 복수의 웹 문서를 검색하는 단계;

각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 단계; 및

상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 단계를 포함하고,

상기 문서 순위 점수를 산출하는 단계는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하는 단계를 포함하고,

상기 복수의 모수 중 적어도 하나를 업데이트하는 단계는 상기 복수의 웹 문서의 과거 문서 순위 점수와 검색 만족도의 오차를 이용하여 상기 복수의 모수 중 적어도 하나를 업데이트하며,

상기 복수의 모수 중 적어도 하나를 업데이트하는 단계는 복수의 사용자를 포함하는 사용자 그룹 또는 복수의 검색어를 포함하는 검색어 그룹에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하는, 검색서버에서 수행되는 검색 방법.

청구항 3

사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 단계;

상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에 관련된 복수의 웹 문서를 검색하는 단계;

각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 단계; 및

상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 단계를 포함하고,

상기 문서 순위 점수를 산출하는 단계는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하는 단계를 포함하고,

상기 복수의 모수 중 적어도 하나를 업데이트하는 단계는 상기 복수의 웹 문서의 과거 문서 순위 점수와 검색 만족도의 오차를 이용하여 상기 복수의 모수 중 적어도 하나를 업데이트하며,

상기 복수의 모수 중 적어도 하나를 업데이트하는 단계는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 특징 중 검색 만족도가 기준 값보다 높거나 낮은 특징을 판단하고, 상기 판단된 특징에 대응하는

모수를 상향 조정하거나 하향 조정하는, 검색서버에서 수행되는 검색 방법.

청구항 4

제3항에 있어서,

상기 검색 만족도는 CTR(Click-Through Rate), VTR(View-Through Rate) 중 적어도 하나를 포함하는, 검색 방법.

청구항 5

사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 단계;

상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에 관련된 복수의 웹 문서를 검색하는 단계;

각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 단계; 및

상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 단계를 포함하고,

상기 문서 순위 점수를 산출하는 단계는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하는 단계를 포함하고,

상기 복수의 모수 중 적어도 하나를 업데이트하는 단계는 상기 복수의 웹 문서의 과거 문서 순위 점수와 검색 만족도의 오차를 이용하여 상기 복수의 모수 중 적어도 하나를 업데이트하며,

상기 복수의 특징은 최신성(freshness), 유사성(similarity), 품질(quality), 검색 로그(search log) 중 적어도 하나를 포함하는, 검색서버에서 수행되는 검색 방법.

청구항 6

사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 단계;

상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에 관련된 복수의 웹 문서를 검색하는 단계;

각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 단계; 및

상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 단계를 포함하고,

상기 문서 순위 점수를 산출하는 단계는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하는 단계를 포함하고,

상기 복수의 모수 중 적어도 하나를 업데이트하는 단계는 상기 복수의 웹 문서의 과거 문서 순위 점수와 검색 만족도의 오차를 이용하여 상기 복수의 모수 중 적어도 하나를 업데이트하며,

상기 복수의 웹 문서의 문서 순위 점수를 산출하는 단계는 Pointwise 접근법, Pairwise 접근법, Listwise 접근법 중 적어도 하나를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는, 검색서버에서 수행되는 검색 방법.

청구항 7

삭제

청구항 8

제1항에 있어서,

상기 업데이트되는 복수의 모수 중 적어도 하나의 변화량은 상기 오차의 그라디언트 및 학습률을 이용하여 결정되는, 검색 방법.

청구항 9

사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 검색 요청 수신부;

상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에 관련된 복수의 웹 문서를 검색하는 정보 검색부;

각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 검색 순위부; 및

상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 검색 결과 제공부를 포함하고,

상기 검색 순위부는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하며,

상기 검색 순위부는 상기 복수의 웹 문서의 과거 문서 순위 점수와 검색 만족도의 오차를 이용하여 상기 복수의 모수 중 적어도 하나를 업데이트하는 것을 특징으로 하는, 검색 서버.

청구항 10

사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 검색 요청 수신부;

상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에 관련된 복수의 웹 문서를 검색하는 정보 검색부;

각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 검색 순위부; 및

상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 검색 결과 제공부를 포함하고,

상기 검색 순위부는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하고,

상기 검색 순위부는 상기 복수의 웹 문서의 과거 문서 순위 점수와 검색 만족도의 오차를 이용하여 상기 복수의 모수 중 적어도 하나를 업데이트하며,

상기 검색 순위부는 복수의 사용자를 포함하는 사용자 그룹 또는 복수의 검색어를 포함하는 검색어 그룹에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하는, 검색 서버.

청구항 11

사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 검색 요청 수신부;

상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에 관련된 복수의 웹 문서를 검색하는 정보 검색부;

각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 검색 순위부; 및

상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 검색 결과 제공부를 포함하고,

상기 검색 순위부는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하고,

상기 검색 순위부는 상기 복수의 웹 문서의 과거 문서 순위 점수와 검색 만족도의 오차를 이용하여 상기 복수의 모수 중 적어도 하나를 업데이트하며,

상기 검색 순위부는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 특징 중 검색 만족도가 기준 값보다 높거나 낮은 특징을 판단하고, 상기 판단된 특징에 대응하는 모수를 상향 조정하거나 하향 조정하는, 검색 서버.

청구항 12

제11항에 있어서,

상기 검색 만족도는 CTR(Click-Through Rate), VTR(View-Through Rate) 중 적어도 하나를 포함하는, 검색 서버.

청구항 13

사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 검색 요청 수신부;

상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에 관련된 복수의 웹 문서를 검색하는 정보 검색부;

각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 검색 순위부; 및

상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 검색 결과 제공부를 포함하고,

상기 검색 순위부는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하고,

상기 검색 순위부는 상기 복수의 웹 문서의 과거 문서 순위 점수와 검색 만족도의 오차를 이용하여 상기 복수의 모수 중 적어도 하나를 업데이트하며,

상기 복수의 특징은 최신성(freshness), 유사성(similarity), 품질(quality), 검색 로그(search log) 중 적어도 하나를 포함하는, 검색 서버.

청구항 14

사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 검색 요청 수신부;

상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에 관련된 복수의 웹 문서를 검색하는 정보 검색부;

각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 검색 순위부; 및

상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 검색 결과 제공부를 포함하고,

상기 검색 순위부는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하고,

상기 검색 순위부는 상기 복수의 웹 문서의 과거 문서 순위 점수와 검색 만족도의 오차를 이용하여 상기 복수의 모수 중 적어도 하나를 업데이트하며,

상기 검색 순위부는 Pointwise 접근법, Pairwise 접근법, Listwise 접근법 중 적어도 하나를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는, 검색 서버.

청구항 15

삭제

청구항 16

제9항에 있어서,

상기 업데이트되는 복수의 모수 중 적어도 하나의 변화량은 상기 오차의 그라디언트 및 학습률을 이용하여 결정되는, 검색 서버.

발명의 설명

기술 분야

[0001] 본 발명은 검색 방법 및 검색 서버에 관한 것이다.

배경 기술

[0002] 검색 엔진은 인터넷에 존재하는 수 많은 웹 문서들을 수집하고, 웹 문서들에 포함된 내용들을 분석하여 색인화(indexing)한다. 사용자는 사용자 단말기를 이용하여 검색어를 포함하는 검색 요청을 검색 엔진에 전송하고, 검색 엔진은 사용자의 검색 요청에 응답하여 검색어와 관련된 복수의 웹 문서들을 검색한다. 검색 엔진은 사전에 색인화된 웹 문서들의 정보를 조회하여 어떤 웹 문서들이 검색어와 매칭되는지를 판단한다. 검색 엔진은 검색어와 매칭되는 복수의 웹 문서들을 일정한 순위에 따라 정렬하여 검색 결과로 제공한다. 이 때, 복수의 웹 문서들

의 순위를 결정하기 위하여, 검색 엔진은 각각의 웹 문서들의 문서 순위 점수를 산출할 수 있다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) 한국등록특허공보 제10-0923505호, 2009년 10월 19일

발명의 내용

해결하려는 과제

[0004] 본 발명이 해결하고자 하는 과제는 사용자의 검색 요청에 응답하여 사용자에게 보다 정확한 검색 결과를 제공하는 검색 방법 및 검색 서버를 제공하는 것이다.

[0005] 본 발명이 해결하고자 하는 다른 과제는 사용자의 검색 요청에 응답하여 사용자에게 사용자 만족도가 보다 높은 검색 결과를 제공하는 검색 방법 및 검색 서버를 제공하는 것이다.

[0006] 본 발명이 해결하고자 하는 과제는 이상에서 언급한 과제로 제한되지 않으며, 언급하지 않은 또 다른 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0007] 상기 과제를 해결하기 위한 본 발명의 일 측면(Aspect)에 따른 검색 방법은 사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 단계, 상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에 관련된 복수의 웹 문서를 검색하는 단계, 각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 단계, 및 상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 단계를 포함하고, 상기 문서 순위 점수를 산출하는 단계는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하는 단계를 포함하는 것을 특징으로 한다.

[0008] 본 발명의 몇몇 실시예에서, 상기 복수의 모수 중 적어도 하나를 업데이트하는 단계는 복수의 사용자를 포함하는 사용자 그룹 또는 복수의 검색어를 포함하는 검색어 그룹에 따라 상기 복수의 모수 중 적어도 하나를 업데이트할 수 있다.

[0009] 본 발명의 몇몇 실시예에서, 상기 복수의 모수 중 적어도 하나를 업데이트하는 단계는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 특징 중 검색 만족도가 기준 값보다 높거나 낮은 특징을 판단하고, 상기 판단된 특징에 대응하는 모수를 상향 조정하거나 하향 조정할 수 있다.

[0010] 또한, 상기 검색 만족도는 CTR(Click-Through Rate), VTR(View-Through Rate) 중 적어도 하나를 포함할 수 있다.

[0011] 본 발명의 몇몇 실시예에서, 상기 복수의 특징은 최신성(freshness), 유사성(similarity), 품질(quality), 검색 로그(search log) 중 적어도 하나를 포함할 수 있다.

[0012] 본 발명의 몇몇 실시예에서, 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 단계는 Pointwise 접근법, Pairwise 접근법, Listwise 접근법 중 적어도 하나를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출할 수 있다.

[0013] 본 발명의 몇몇 실시예에서, 상기 복수의 모수 중 적어도 하나를 업데이트하는 단계는 상기 복수의 웹 문서의 과거 문서 순위 점수와 검색 만족도의 오차를 이용하여 상기 복수의 모수 중 적어도 하나를 업데이트할 수 있다.

[0014] 또한, 상기 업데이트되는 복수의 모수 중 적어도 하나의 변화량은 상기 오차의 그라디언트 및 학습률을 이용하여 결정될 수 있다.

[0015] 상기 과제를 해결하기 위한 본 발명의 일 측면에 따른 검색 서버는 사용자 단말기로부터 적어도 하나의 검색어를 포함하는 검색 요청을 수신하는 검색 요청 수신부, 상기 검색 요청에 응답하여 상기 적어도 하나의 검색어에

관련된 복수의 웹 문서를 검색하는 정보 검색부, 각각의 웹 문서의 복수의 특징(feature) 및 상기 복수의 특징에 각각 대응하는 복수의 모수(parameter)를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 검색 순위부, 및 상기 문서 순위 점수에 따라 상기 사용자 단말기에 검색 결과를 제공하는 검색 결과 제공부를 포함하고, 상기 검색 순위부는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 모수 중 적어도 하나를 업데이트하여 상기 복수의 웹 문서의 문서 순위 점수를 산출하는 것을 특징으로 한다.

- [0016] 본 발명의 몇몇 실시예에서, 상기 검색 순위부는 복수의 사용자를 포함하는 사용자 그룹 또는 복수의 검색어를 포함하는 검색어 그룹에 따라 상기 복수의 모수 중 적어도 하나를 업데이트할 수 있다.
- [0017] 본 발명의 몇몇 실시예에서, 상기 검색 순위부는 상기 사용자 또는 상기 적어도 하나의 검색어에 따라 상기 복수의 특징 중 검색 만족도가 기준 값보다 높거나 낮은 특징을 판단하고, 상기 판단된 특징에 대응하는 모수를 상향 조정하거나 하향 조정할 수 있다.
- [0018] 또한, 상기 검색 만족도는 CTR(Click-Through Rate), VTR(View-Through Rate) 중 적어도 하나를 포함할 수 있다.
- [0019] 본 발명의 몇몇 실시예에서, 상기 복수의 특징은 최신성(freshness), 유사성(similarity), 품질(quality), 검색 로그(search log) 중 적어도 하나를 포함할 수 있다.
- [0020] 본 발명의 몇몇 실시예에서, 상기 순위 결정부는 Pointwise 접근법, Pairwise 접근법, Listwise 접근법 중 적어도 하나를 이용하여 상기 복수의 웹 문서의 문서 순위 점수를 산출할 수 있다.
- [0021] 본 발명의 몇몇 실시예에서, 상기 검색 순위부는 상기 복수의 웹 문서의 과거 문서 순위 점수와 검색 만족도의 오차를 이용하여 상기 복수의 모수 중 적어도 하나를 업데이트할 수 있다.
- [0022] 또한, 상기 업데이트되는 복수의 모수 중 적어도 하나의 변화량은 상기 오차의 그라디언트 및 학습률을 이용하여 결정될 수 있다.
- [0023] 본 발명의 기타 구체적인 사항들은 상세한 설명 및 도면들에 포함되어 있다.

발명의 효과

- [0024] 상기 본 발명의 검색 방법 및 검색 서버에 의하면, 문서 순위 점수를 산출하기 위하여 이용되는 모수를 고정시키지 않고 사용자 또는 검색어에 따라 업데이트하므로, 사용자에게 보다 정확하고 사용자 만족도가 보다 높은 검색 결과를 제공할 수 있다.
- [0025] 본 발명의 효과는 이상에서 언급한 효과로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [0026] 도 1은 본 발명의 일 실시예에 따른 검색 서버가 제공되는 환경을 설명하기 위한 도면이다.
- 도 2는 도 1의 검색 서버의 세부 구성을 설명하기 위한 도면이다.
- 도 3은 본 발명의 일 실시예에 따른 검색 방법을 설명하기 위한 흐름도이다.
- 도 4는 도 3의 문서 순위 점수 결정 단계를 세부적으로 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0027] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나, 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 것이며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하며, 통상의 기술자에게 본 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.
- [0028] 하나의 구성요소(elements)가 다른 구성요소와 "연결된(connected to)"이라고 지칭되는 것은, 다른 구성요소와 직접 연결 또는 중간에 다른 구성요소를 개재한 경우를 모두 포함한다. 반면, 하나의 구성요소가 다른 구성요소와 "직접 연결된(directly connected to)"으로 지칭되는 것은 중간에 다른 구성요소를 개재하지 않은 것을 나타

낸다.

- [0029] 본 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성요소를 지칭한다.
- [0030] 본 명세서에서 "및/또는"은 언급된 구성요소의 각각 및 하나 이상의 모든 조합을 포함한다.
- [0031] 비록 제1, 제2 등이 다양한 구성요소, 모듈 및/또는 섹션들을 서술하기 위해서 사용되나, 구성요소, 모듈 및/또는 섹션들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 구성요소, 모듈 또는 섹션들을 다른 구성요소, 모듈 또는 섹션들과 구별하기 위하여 사용하는 것이다. 따라서, 이하에서 언급되는 제1 구성요소, 제1 모듈 또는 제1 섹션은 본 발명의 기술적 사상 내에서 제2 구성요소, 제2 모듈 또는 제2 섹션일 수도 있음은 물론이다.
- [0032] 본 명세서에서 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다.
- [0033] 본 명세서에서 사용되는 "포함한다(comprises)" 및/또는 "포함하는(comprising)"은 언급된 구성요소, 모듈 및/또는 섹션 외에 하나 이상의 다른 구성요소, 모듈 및/또는 섹션의 존재 또는 추가를 배제하지 않는다.
- [0034] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 통상의 기술자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또한, 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다.
- [0035] 도 1은 본 발명의 일 실시예에 따른 검색 서버가 제공되는 환경을 설명하기 위한 도면이다.
- [0036] 도 1을 참조하면, 복수의 사용자 단말기(100)가 검색 서버(200)와 네트워크를 통해 서로 연결된다. 사용자 단말기(100)와 검색 서버(200)는 서로 데이터 및/또는 정보를 송수신할 수 있다.
- [0037] 네트워크는 근거리 네트워크(Local Area Network; LAN), 도시권 네트워크(Metropolitan Area Network; MAN), 광대역 네트워크(Wide Area Network; WAN) 등과 같은 다양한 크기의 네트워크로 구성될 수 있다. 네트워크는 유선 또는 무선 네트워크로 구성될 수 있다.
- [0038] 사용자 단말기(100)는 데스크톱(desk top), 랩톱(lap top) 등과 같은 개인용 컴퓨터(Personal Computer; PC)일 수 있다. 또는, 사용자 단말기(100)는 스마트폰(smartphone), PDA(Personal Digital Assistant), 태블릿 PC(tablet PC) 등과 같은 휴대용 통신 장치일 수 있다. 사용자 단말기(100)는 프로세서, 입출력 수단, 통신 수단을 포함하는 예시되지 않은 다른 컴퓨팅 장치일 수도 있다.
- [0039] 검색 서버(200)는 웹 문서의 수집, 처리 및 검색 기능을 제공하는 검색 엔진을 포함할 수 있다. 즉, 검색 서버(200)는 인터넷에 존재하는 웹 문서들을 크롤링(crawling), 색인화(indexing), 검색(searching)하여, 웹 문서들의 정보를 데이터베이스에 저장할 수 있다.
- [0040] 본 명세서에서 “웹 문서”는 임의의 컴퓨팅 장치로 관독 및 저장 가능한 객체를 나타낸다. 예를 들어, 웹 문서는 텍스트뿐만 아니라, 이미지, 소리, 동영상 등을 포함할 수 있다. 또한, 웹 문서는 스크립트와 같은 명령어들을 포함할 수도 있다. 웹 문서에는 다른 웹 문서에 대한 링크가 삽입될 수 있다. 본 명세서에서 웹 문서는 “웹 페이지”로 대체될 수도 있다.
- [0041] 검색 서버(200)는 사용자 단말기(100)로부터 검색 요청을 수신하고, 검색 요청에 응답하여 검색어에 관련된 복수의 웹 문서들을 검색하고, 검색된 복수의 웹 문서들의 문서 순위 점수를 산출하고, 문서 순위 점수에 따라 사용자 단말기(100)에 검색 결과를 제공할 수 있다.
- [0042] 검색 서버(200)는 검색 엔진의 기능 외에 카페, 메일, 블로그, 쇼핑, 지도, 사진, 뉴스, 증권, 부동산, 영화, 음악, 게시판 등의 다양한 콘텐츠 서비스를 제공하는 포털 사이트 서버일 수 있다. 즉, 검색 서버(200)는 검색 엔진과 포털 사이트가 결합된 형태일 수 있다.
- [0043] 도 1에서는 하나의 검색 서버(200)만을 도시하였으나, 본 발명이 이에 한정되는 것은 아니고, 복수의 검색 서버(200)가 사용자 단말기(100)와 네트워크를 통해 서로 연결될 수 있음은 통상의 기술자에게 자명할 것이다. 예를 들어, 크롤링하는 서버, 색인화하는 서버, 검색하는 서버가 각각 하나의 서버로 구성될 수 있다.
- [0044] 본 발명의 실시예에서, 검색 서버(200)는 문서 순위 점수를 산출하기 위하여 이용되는 모수를 고정시키지 않고 사용자 또는 검색어에 따라 업데이트할 수 있다. 이로써, 검색 서버(200)는 사용자에게 보다 정확하고 사용자 만족도가 높은 검색 결과를 제공할 수 있다. 모수의 업데이트 방법에 관하여는 이하에서 보다 상세하게 설명하기로 한다.

- [0045] 도 2는 도 1의 검색 서버의 세부 구성을 설명하기 위한 도면이다.
- [0046] 도 2를 참조하면, 도 1의 검색 서버(200)는 검색 요청 수신부(210), 정보 검색부(220), 검색 순위부(230), 검색 결과 제공부(240), 문서 데이터베이스(250), 학습 데이터(260)을 포함한다.
- [0047] 검색 요청 수신부(210)는 사용자 단말기(100)로부터 검색어를 포함하는 검색 요청을 수신한다. 사용자의 검색 요청은 복수의 검색어를 포함할 수 있다.
- [0048] 정보 검색부(220)는 전체 웹 문서 중 검색어에 관련된(또는, 매칭되는) 복수의 웹 문서를 검색한다. 정보 검색부(220)는 문서 데이터베이스(250)에 저장된 웹 문서들의 정보를 조회하여 어떤 웹 문서들이 검색어와 관련되는지를 판단할 수 있다. 이를 위하여, 정보 검색부(220)는 역색인화(reverse-indexing)를 이용할 수 있다.
- [0049] 검색 순위부(230)는 검색된 복수의 웹 문서의 문서 순위 점수를 산출한다. 문서 순위 점수는 각각의 웹 문서가 검색어와 얼마나 관련되는지를 나타낸다.
- [0050] 문서 순위 점수를 산출하기 위하여, 검색 순위부(230)는 각각의 웹 문서의 복수의 특징(feature)을 이용할 수 있다. 본 명세서에서, 웹 문서의 “특징”은 문서 순위 점수를 산출하기 위하여 산출 방법(또는, 공식)에서 이용되는 요소(element)를 나타낸다. 예를 들어, 복수의 특징은 최신성(freshness), 유사성(similarity), 품질(quality), 검색 로그(search log) 등을 포함할 수 있다.
- [0051] 실시예에 따라, 복수의 특징은 각각의 웹 문서 자체에서 유래된 요소, 각각의 웹 문서와 검색어와의 관계에서 유래된 요소, 각각의 웹 문서와 사용자와의 관계에서 유래된 요소, 상술한 요소들의 조합에 의해 유래된 요소 등으로 구성될 수 있다.
- [0052] 복수의 특징은 예시되지 않은 다른 요소를 포함할 수도 있다.
- [0053] 또한, 문서 순위 점수를 산출하기 위하여, 검색 순위부(230)는 복수의 모수(parameter)를 이용할 수 있다. 본 명세서에서, “모수”는 웹 문서의 문서 순위 점수 산출 방법에서 각각의 특징에 부여되는 가중치를 나타낸다. 복수의 모수는 상술한 복수의 특징에 각각 대응할 수 있다.
- [0054] 검색 순위부(230)는 Pointwise 접근법, Pairwise 접근법, Listwise 접근법 등을 이용하여 문서 순위 점수를 산출할 수 있다. 이하에서는, 설명의 편의를 위하여, Pointwise 접근법 중 Regression 방법을 이용하는 것을 예시로 하여 설명기로 한다.
- [0055] 검색 순위부(230)는 아래 수학적 식 1에 따라 문서 순위 점수를 산출할 수 있다.
- [0056] [수학적 식 1]

$$H(x) = ax + bx' + cx''$$

- [0057]
- [0058] 수학적 식 1에서, H(x)는 문서 순위 점수를 나타내고, x, x', x''는 특징을 나타내고, a, b, c는 모수를 나타낸다.
- [0059] x는 최신성 확률이고, x'는 유사 비율이고, x''는 품질 점수라고 가정하자. 최신성 확률은 웹 문서가 얼마나 최근에 생성된 것인지를 나타내는 지표이고, 유사 비율은 웹 문서에 포함된 내용들이 검색어와 얼마나 관련되는지를 나타내는 지표이고, 품질 점수는 웹 문서의 생성자, 출처 등을 평가한 점수를 나타낸다. 품질 점수가 이용되는 것은 스팸머(spamer)에 의해 생성된 웹 문서를 필터링하기 위함이다.
- [0060] 모수 a, b, c는 학습 데이터를 이용하여 산출될 수 있다. 예를 들어, 사용자가 뉴스 문서를 검색하는 경우, 최신성 확률과 유사 비율이 품질 점수에 비하여 상대적으로 중요하므로, 모수 a, b, c는 a>b>c와 같이 산출될 수 있다. 다른 예를 들어, 사용자가 블로그 문서를 검색하는 경우, 품질 점수가 최신성 확률 및 유사 비율에 비하여 상대적으로 중요하게 고려될 수 있으므로, 모수 a, b, c는 a<b<c와 같이 산출될 수 있다.
- [0061] 본 예시에서, a, b, c는 각각 0.3, 0.4, 0.5라고 가정하자.
- [0062] 정보 검색부(220)가 검색어와 관련된 문서로 두 개의 문서(문서 1 및 문서 2)를 검색하였다고 가정하자. 문서 1은 x=0.5, x'=0.7, x''=0.8의 특징 값을 가진다고 가정하고, 문서 2는 x=0.8, x'=0.7, x''=0.5의 특징 값을 가진다고 가정하면, 문서 1 및 문서 2의 문서 순위 점수는 다음과 같다.

- [0063] 문서 1의 문서 순위 점수:
- [0064] $0.3 \times 0.5 + 0.4 \times 0.7 + 0.5 \times 0.8 = 0.83$
- [0065] 문서 2의 문서 순위 점수:
- [0066] $0.3 \times 0.8 + 0.4 \times 0.7 + 0.5 \times 0.5 = 0.77$
- [0067] 여기서, 모수가 $a=0.3$, $b=0.4$, $c=0.5$ 로 계속 고정된다고 하면, 시간이 경과하거나 또는 모수를 산출했던 학습 데이터와 다른 양상이 나타나는 경우 검색 정확도가 낮아지게 되는 문제점이 발생하게 된다.
- [0068] 이 같은 문제점을 해결하기 위하여, 본 발명의 검색 순위부(230)는 모수의 우도(likelihood)값을 산출하여 모수를 업데이트한다. 즉, 검색 순위부(230)는 문서 순위 점수를 산출하는 때에 가변적인 모수를 이용한다.
- [0069] 검색 순위부(230)는 복수의 모수 전부를 업데이트하거나, 복수의 모수 중 일부를 업데이트할 수 있다.
- [0070] 검색 순위부(230)는 사용자 또는 검색어에 따라 모수를 업데이트할 수 있다.
- [0071] 예를 들어, 검색어에 따른 모수의 업데이트값이 최신성 확률의 경우 2, 유사 비율의 경우 1, 품질 점수의 경우 0.5로 산출되었다고 가정하고, 검색 요청을 진송한 사용자에 따른 모수의 업데이트 값이 최신성 확률의 경우 3, 유사 비율의 경우 2, 품질 점수의 경우 1로 산출되었다고 가정하자.
- [0072] 위의 모수의 업데이트값을 고려하여 문서 1 및 문서 2의 문서 순위 점수를 산출하면 다음과 같다.
- [0073] 문서 1의 문서 순위 점수:
- [0074] $0.3 \times 2 \times 3 \times 0.5 + 0.4 \times 1 \times 2 \times 0.7 + 0.5 \times 0.5 \times 1 \times 0.8 = 1.66$
- [0075] 문서 2의 문서 순위 점수:
- [0076] $0.3 \times 2 \times 3 \times 0.8 + 0.4 \times 1 \times 2 \times 0.7 + 0.5 \times 0.5 \times 1 \times 0.5 = 2.125$
- [0077] 실시예에 따라, 업데이트값은 음의 값을 가질 수도 있다.
- [0078] 검색어가 복수 개 존재하는 경우, 일부 검색어 또는 검색어 전부가 고려될 수 있다. 이 경우, 검색어에 따른 모수의 업데이트값은 각각의 검색어에 대응하여 존재하거나, 일부 검색어에 대응하여 존재할 수 있다. 검색어에 따른 모수의 업데이트값은 검색어 전부를 고려하여 하나만 존재할 수도 있다.
- [0079] 위의 예시에서, 검색 순위부(230)는 현재의 모수에 검색어에 따른 모수의 업데이트값과 사용자에 따른 모수의 업데이트값을 각각 이용하였으나, 실시예에 따라, 현재의 모수에 검색어 및 사용자를 고려한 하나의 모수의 업데이트값을 이용할 수도 있다.
- [0080] 이상의 예시에서는 현재의(또는, 디폴트(default)) 모수에 모수의 업데이트값을 곱하여 모수를 업데이트하였으나, 본 발명이 이에 한정되는 것은 아니고, 현재의 모수에 모수의 변화량을 더하거나 빼서 모수를 업데이트할 수 있다. 또는, 현재의 모수를 요소로 하는 함수를 이용하여 모수를 업데이트할 수도 있다.
- [0081] 이상에서 설명한 모수의 업데이트를 위하여, 검색 순위부(230)는 검색 만족도를 이용할 수 있다. 검색 만족도는 검색 로그를 분석하여 획득될 수 있다. 검색 로그는 사용자 단말기(100)가 검색 서버(200)에 접속하여 요청, 처리 또는 이용하는 과정에서 발생한 일련의 정보가 기록된 자료를 나타낸다. 이와 같은 검색 로그는 어떤 사용자가 접속했는지, 어떤 검색어를 입력하였는지, 어떤 웹 문서를 클릭하거나 열람하였는지, 웹 브라우저의 종류는 무엇인지 등의 정보를 포함할 수 있다. 검색 순위부(230)는 이와 같은 검색 로그를 분석하여 획득되는 CTR(Click-Through Rate) 또는 VTR(View-Through Rate) 등의 검색 만족도를 이용할 수 있다.
- [0082] 검색 순위부(230)는 사용자별 검색 만족도 또는 검색어별 검색 만족도를 고려하여 소정의 모수를 상향 조정할 수 있다. 검색 순위부(230)는 웹 문서의 복수의 특징 중 검색 만족도가 기준 값(reference value)보다 높은 특징을 판단하고, 해당하는 특징에 대응하는 모수를 상향 조정할 수 있다.
- [0083] 이를 위하여, 웹 문서의 각각의 특징의 검색 만족도를 산출하기 위한 다양한 방법들이 이용될 수 있을 것이다. 예를 들어, 복수의 웹 문서가 검색 결과에 포함되어 제공될 때, 어느 웹 문서는 다른 특징들보다 최신성 확률이 상대적으로 높다는 이유로 제공되고, 다른 웹 문서는 다른 특징들보다 유사 비율이 상대적으로 높다는 이유로 제공될 수 있다. 위의 예시에서, 각각의 웹 문서의 검색 만족도(CTR 등)는 서로 다르게 측정될 것이고, 이를 이용하면 각각의 특징의 검색 만족도가 산출될 수 있을 것이다.

[0084] 한편, 기준 값은 복수의 특징의 검색 만족도의 통계에 따른 값(예를 들어, 평균 값 또는 중간 값 등)으로 사전 설정되거나, 검색 서버(200)의 운영자가 입력한 임의의 값으로 사전 설정될 수 있다.

[0085] 위의 예시에서, 특정 사용자와 관련하여 품질 점수의 검색 만족도가 기준 값보다 높다고 가정하면, 검색 순위부(230)는 품질 점수에 대응하는 모수 c를 상향조정할 수 있다. 또는, 특정 검색어와 관련하여 최신성 확률의 검색 만족도가 기준 값보다 높다고 가정하면, 검색 순위부(230)는 최신성 확률에 대응하는 모수 a를 상향조정할 수 있다.

[0086] 또한, 검색 순위부(230)는 사용자별 검색 만족도 또는 검색어별 검색 만족도를 고려하여 소정의 모수를 하향 조정할 수 있다. 검색 순위부(230)는 웹 문서의 복수의 특징 중 검색 만족도가 기준 값보다 낮은 특징을 판단하고, 해당하는 특징에 대응하는 모수를 하향 조정할 수 있다.

[0087] 실시예에 따라, 검색 만족도를 (기준 값이 아닌) 복수의 레벨로 구분하고, 각각의 레벨에 따라 모수를 조정할 수도 있다. 또는, 복수의 특징 중 검색 만족도가 가장 높거나 가장 낮은 특징에 대응하는 모수만을 조정할 수도 있다.

[0088] 한편, (예를 들어, 사용자의 검색 로그의 데이터량이 너무 적거나, 또는 분석 결과에 따른 모수의 업데이트값이 너무 작거나 하는 등의 이유로) 사용자의 검색 만족도 분석 결과가 무의미하다면, 검색 순위부(230)는 해당 사용자와 유사한 다른 사용자들의 검색 만족도 분석 결과를 이용할 수도 있다. 즉, 검색 순위부(230)는 복수의 사용자를 포함하는 사용자 그룹의 검색 만족도를 이용할 수 있다. 하나의 사용자 그룹에 포함되는 복수의 사용자는 소정의 유사도를 가질 수 있다. 사용자 간의 유사도는 사용자의 프로파일 정보(연령, 성별, 주소, 학력, 직업, 결혼 여부 등) 또는 사용자의 행동 정보(입력 검색어, 브라우징 활동에 관한 정보, 구매 정보 등) 등을 이용하여 산출될 수 있다.

[0089] 또한, 검색 순위부(230)는 복수의 검색어를 포함하는 검색어 그룹의 검색 만족도를 이용할 수도 있다. 하나의 검색어 그룹에 포함되는 복수의 검색어는 소정의 유사도를 가질 수 있다. 검색어 간의 유사도는 입력 사용자, 입력 시점, 편집 거리, 검색 결과의 유사성 등을 이용하여 산출될 수 있다.

[0090] 이로써, 검색 순위부(230)의 모수의 업데이트 방법은 (지나치게 치우치지 않는) 적절한 학습률(learning rate)을 찾고, 재현율(recall)을 증가시킬 수 있다.

[0091] 검색 만족도를 이용하여 모수를 업데이트하기 위하여, 검색 순위부(230)는 역전파(backpropagation) 알고리즘과 유사한 알고리즘을 이용할 수 있다. 역전파 알고리즘은 오차가 최소가 되도록 가중치를 조정하는 방법으로, 본 발명이 속하는 기술분야에서 잘 알려진 다른 알고리즘이 이용될 수도 있을 것이다.

[0092] 검색 순위부(230)는 아래 수학식 2에 따라 추정된 검색 만족도와 실제 검색 만족도의 평균제곱오차(MSE; Mean Square Error)를 산출할 수 있다.

[0093] [수학식 2]

$$C = \frac{1}{2} \sum (H(x) - y)^2 = \frac{1}{2} \sum ((w^T x + b) - y)^2$$

[0094]

[0095] 수학식 2에서, C는 오차를 나타내고, H(x)는 과거 문서 순위 점수(즉, 공식에 의해 추정된 검색 만족도), y는 측정된 검색 만족도(즉, 실제 만족도)를 나타낸다. 그리고, w는 모수의 행렬, x는 특징의 행렬을 나타내고, b는 대응하는 모수가 존재하지 않는 바이패스(bypass) 특징을 나타낸다.

[0096] 모수의 행렬 w의 상태는 아래 수학식 3과 같이 정의될 수 있다.

[0097] [수학식 3]

$$w(h + 1) = w(h) + \Delta w$$

[0098]

[0099] 수학식 3에서, w(h+1)은 다음의(업데이트된) 모수의 행렬, w(h)는 현재의 모수의 행렬, Δw는 모수의 변화량을 나타낸다.

[0100] 검색 순위부(230)는 오차의 그라디언트(gradient)를 이용하여 모수의 변화량을 결정할 수 있다. 이 때, 오차의 그라디언트는 오차의 크기가 작아지는 방향으로 이동하므로, 검색 순위부(230)는 현재의 모수로부터 오차의 그라디언트를 빼서 오차 C가 작아지게 할 수 있다. 그리고, 모수를 얼마나 변화시켜야 오차 C가 작아지는지 불명확하므로, 학습률(learning rate) r이 오차의 그라디언트에 곱해질 수 있다. 학습률은 모수의 변화량을 조절하기 위하여 사전 설정된 상수(constant) 비율을 나타내며, 오차의 그라디언트의 변화량이 모수의 변화량에 미치는 영향을 제어하도록 기능한다. 예를 들어, 학습률은 0 내지 1 사이의 값을 가질 수 있다. 학습률의 산출 방법은 기계 학습 또는 인공지능의 기술분야에서 잘 알려진 다양한 알고리즘이 이용될 수 있을 것이다. 이와 같이 모수의 변화량을 결정하는 방법은 아래 수학적 식 4와 같이 표현될 수 있다.

[0101] [수학적 식 4]

$$w(h + 1) = w(h) + r \frac{\partial C}{\partial x}$$

[0102]

[0103] 검색 결과 제공부(240)는 산출된 문서 순위 점수에 따라 사용자 단말기(100)에 검색 결과를 제공한다. 검색 결과 제공부(240)는 검색어에 관련된 복수의 웹 문서를 문서 순위 점수에 따라 정렬하여 검색 결과로 제공할 수 있다.

[0104] 문서 데이터베이스(250)와 학습 데이터(260)는 데이터베이스의 형태로 구성될 수 있다.

[0105] 도 2에서 검색 서버(200)의 일부 구성요소가 다른 구성요소와 결합되어 하나의 구성요소로 구성될 수 있음은 통상의 기술자에게 자명할 것이다.

[0106] 도 3은 본 발명의 일 실시예에 따른 검색 방법을 설명하기 위한 흐름도이고, 도 4는 도 3의 문서 순위 점수 결정 단계를 세부적으로 설명하기 위한 흐름도이다. 설명의 편의를 위하여 도 1과 중복되는 설명은 생략하기로 한다.

[0107] 도 3을 참조하면, 단계 S310에서, 검색 서버(200)는 사용자 단말기(100)로부터 검색어를 포함하는 검색 요청을 수신한다.

[0108] 이어서, 단계 S320에서, 검색 서버(200)는 전체 웹 문서 중 검색어에 관련된 복수의 웹 문서를 검색한다.

[0109] 이어서, 단계 S330에서, 검색 서버(200)는 검색된 복수의 웹 문서의 문서 순위 점수를 결정한다. 상술한 바와 같이, 문서 순위 점수는 각각의 웹 문서가 검색어와 얼마나 관련되는지를 나타낸다. 검색 서버(200)는 Pointwise 접근법, Pairwise 접근법, Listwise 접근법 등을 이용하여 문서 순위 점수를 산출할 수 있다.

[0110] 도 4를 참조하면, 검색 서버(200)는 문서 순위 점수를 산출하기 위하여, 단계 S331에서, 웹 문서의 복수의 특징(feature)을 획득하고, 단계 S332에서 복수의 특징에 대응하는 복수의 모수를 획득할 수 있다. 상술한 바와 같이, 예를 들어, 복수의 특징은 최신성(freshness), 유사성(similarity), 품질(quality), 검색 로그(search log) 등을 포함할 수 있고, 복수의 모수는 상술한 복수의 특징에 각각 대응할 수 있다.

[0111] 이어서, 단계 S333에서, 검색 서버(200)는 사용자 또는 적어도 하나의 검색어에 따라 복수의 모수 중 적어도 하나를 업데이트할 수 있다.

[0112] 모수의 업데이트를 위하여, 검색 서버(200)는 검색 만족도를 이용할 수 있다. 검색 서버(200)는 검색 로그를 분석하여 검색 만족도를 획득할 수 있다. 검색 서버(200)는 검색 로그를 분석하여 획득되는 CTR(Click-Through Rate) 또는 VTR(View-Through Rate) 등의 검색 만족도를 이용할 수 있다.

[0113] 검색 서버(200)는 사용자별 검색 만족도 또는 검색어별 검색 만족도를 고려하여 소정의 모수를 상향 조정하거나 하향 조정할 수 있다. 검색 서버(200)는 웹 문서의 복수의 특징 중 검색 만족도가 기준 값(reference value)보다 높거나 낮은 특징을 판단하고, 해당하는 특징에 대응하는 모수를 상향 조정하거나 하향 조정할 수 있다.

[0114] 실시예에 따라, 검색 만족도를 (기준 값이 아닌) 복수의 레벨로 구분하고, 각각의 레벨에 따라 모수를 조정할 수도 있다. 또는, 복수의 특징 중 검색 만족도가 가장 높거나 가장 낮은 특징에 대응하는 모수만을 조정할 수도 있다.

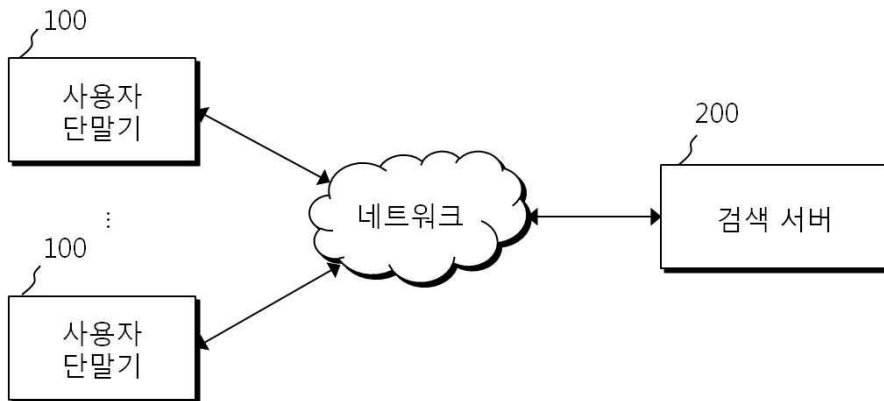
- [0115] 한편, (예를 들어, 사용자의 검색 로그의 데이터량이 너무 적거나, 또는 분석 결과에 따른 모수의 업데이트값이 너무 작거나 하는 등의 이유로) 사용자의 검색 만족도 분석 결과가 무의미하다면, 검색 서버(200)는 해당 사용자와 유사한 다른 사용자들의 검색 만족도 분석 결과를 이용할 수도 있다. 즉, 검색 서버(200)는 복수의 사용자를 포함하는 사용자 그룹의 검색 만족도를 이용할 수 있다. 또한, 검색 서버(200)는 복수의 검색어를 포함하는 검색어 그룹의 검색 만족도를 이용할 수도 있다.
- [0116] 검색 서버(200)는 과거 문서 순위 점수(추정된 검색 만족도)와 측정된 검색 만족도(실제 만족도)의 평균제곱오차(MSE; Mean Square Error)를 산출할 수 있다. 이는 역전파 알고리즘과 유사한 알고리즘을 이용하여 오차가 최소가 되도록 가중치를 조정하기 위한 것이다. 검색 서버(200)는 오차의 그라디언트를 이용하여 모수의 변화량을 결정할 수 있다. 모수를 얼마나 변화시켜야 오차가 작아지는지 불명확하므로, 검색 서버(200)는 학습률(learning rate)을 오차의 그라디언트에 곱하여 모수의 변화량을 결정할 수 있다. 이상에서 설명한 모수의 변화량의 결정 방법은 수학적 2 내지 4를 이용하여 상술한 바와 같다.
- [0117] 이어서, 단계 S334에서, 검색 서버(200)는 복수의 특징 및 업데이트된 복수의 모수를 이용하여 문서 순위 점수를 산출할 수 있다.
- [0118] 이어서, 다시 도 3을 참조하면, 단계 S340에서, 검색 서버(200)는 산출된 문서 순위 점수에 따라 사용자 단말기(100)에 검색 결과를 제공한다. 검색 서버(200)는 검색어에 관련된 복수의 웹 문서를 문서 순위 점수에 따라 정렬하여 검색 결과로 제공할 수 있다.
- [0119] 이상에서 본 발명의 실시예들과 관련하여 설명된 방법 또는 알고리즘의 단계는, 프로세서에 의해 실행되는 하드웨어 모듈, 소프트웨어 모듈, 또는 그 2 개의 결합으로 직접 구현될 수 있다. 소프트웨어 모듈은 RAM(Random Access Memory), ROM, EPROM, EEPROM, 플래시 메모리, 레지스터, 하드 디스크, 착탈형 디스크, CD-ROM, 또는 본 발명이 속하는 기술 분야에서 잘 알려진 임의의 형태의 컴퓨터로 읽을 수 있는 기록 매체에 상주할 수도 있다. 예시적인 기록 매체는 프로세서에 연결되며, 상기 프로세서는 기록 매체로부터 정보를 독출할 수 있고 기록 매체에 정보를 기입할 수 있다. 다른 방법으로, 기록 매체는 프로세서와 일체형일 수도 있다. 프로세서 및 기록 매체는 주문형 집적회로(ASIC) 내에 상주할 수도 있다. ASIC는 사용자 단말기 내에 상주할 수도 있다. 다른 방법으로, 프로세서 및 기록 매체는 사용자 단말기 내에 개별 구성 요소로서 상주할 수도 있다.
- [0120] 이상 첨부된 도면을 참조하여 본 발명의 실시예를 설명하였지만, 통상의 기술자는 본 발명이 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

부호의 설명

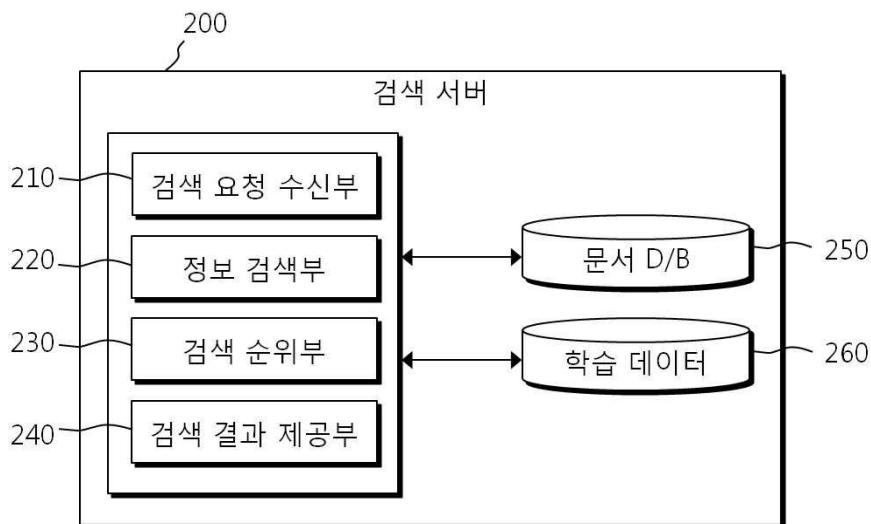
- [0121] 100: 사용자 단말기
- 200: 검색 서버
- 210: 검색 요청 수신부
- 220: 정보 검색부
- 230: 검색 순위부
- 240: 검색 결과 제공부
- 250: 문서 데이터베이스
- 260: 학습 데이터

도면

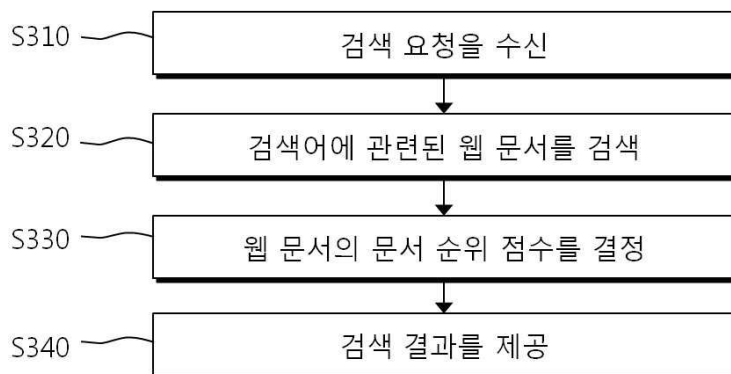
도면1



도면2



도면3



도면4

